Understanding MRI results of patients with/without dementia

1. Project overview

The dataset comes from Kaggle dataset. This dataset consists of a longitudinal collection of nearly 150 subjects aged 60 to 96. Each subject was scanned on two or more visits. We want to compare the MRI results of patients with/without dementia.

Our exploration will address two research questions, which is,

1. What is the impact of dementia on eTIV, nWBV and ASF while accounting for patients' two visits?
2. What is the relationship between age and MMSE and CDR?

2. Data cleaning, preprocessing

A. Dataset Overview

The dataset contains 16 columns, some of them are categorical variables such as Visit, M/F, Hand, some are numerical variables like MMSE, eTiV and so on.

B. Feature Engineering

To ensure the quality and availability of data, we do data cleaning and preprocessing including the following parts:

- Delete unnecessary columns such as 'MRI ID' and 'Hand'.
- Delete some rows with missing values.
- Divide age into 3 age groups, divide MMSE score into 3 groups.

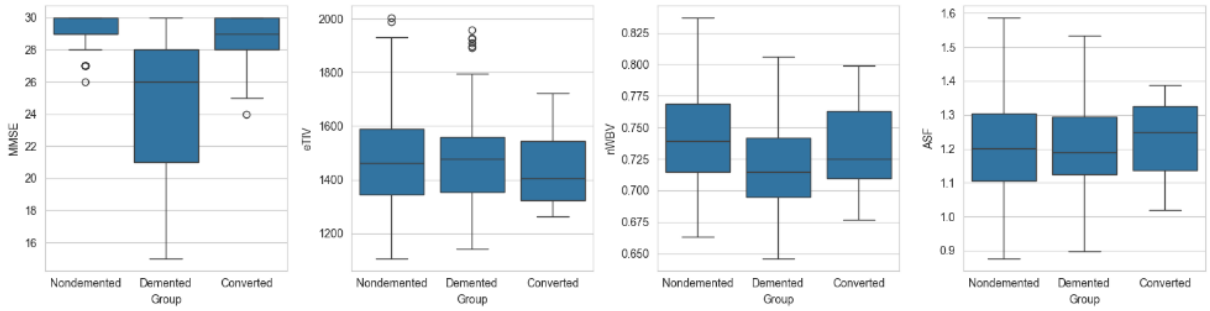
3. Exploratory Data Analysis

First, we start by describing our quantitative data analysis and boxplots to have a visualization on these numerical variables. These can give us a better understanding of these data.

	Age	EDUC	MMSE	eTIV	nWBV	ASF
mean	76.38	14.72	27.28	1479.22	0.73	1.2
std	7.79	2.91	3.45	177.18	0.03	0.14
min	60	6	15	1106	0.64	0.87
max	98	23	30	2004	0.83	1.58

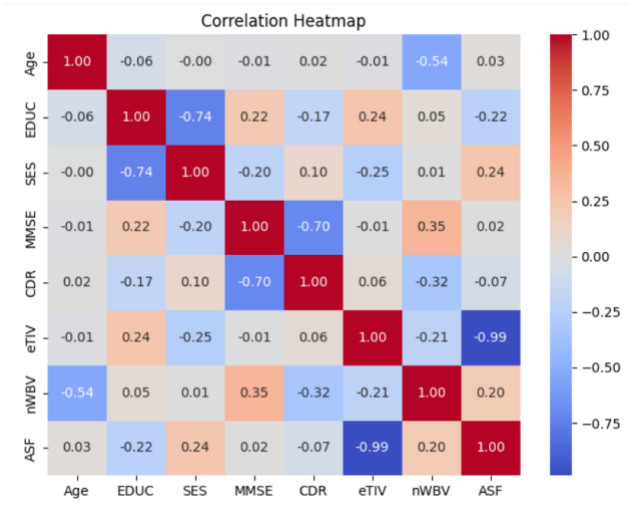
Based on the provided dataset, a statistical analysis reveals several key insights. The average age of the subjects is approximately 76.39 years old, with a standard deviation of 7.79 years, indicating a diverse age range from 60 to 98 years. Education levels reflect a relatively high educational attainment across the sample. Brain structure metrics such as eTIV, nWBV, and ASF demonstrate stable average values, providing insights into the participants' neuroanatomy.

We created boxplots by grouping the data based on patients whether have dementia to explore any potential insights related to brain structures and recognition scores.



From the boxplots above, we can see that patients with dementia tend to a lower MMSE scores. Moreover, for brain structure metrics such as eTIV, nWBV, and ASF, we don't see a large difference between different groups, but we would like to make further analysis in the next part.

Additionally, we want to explore the correlation between some numeric variables, so we plot the correlation map below.



These correlation results provide insights into the interrelationships among age, brain structure (nWBV, eTIV, ASF), cognitive function (MMSE, CDR), and socio-demographic factors (EDUC, SES). The correlation between age and nWBV suggests that as age increases, the normalized whole brain volume tends to decrease, indicating a potential relationship between aging and changes in brain volume. Additionally, the highly positive correlation between eTIV and ASF suggests a strong association between these two brain structure metrics.

4. Impact of dementia

Research Question 1: What is the impact of dementia on eTIV, nWBV and ASF while accounting for patients' two visits? For this research question, we perform mix effects ANOVA to explore the data.

> In statistics, a mixed-design analysis of variance (ANOVA), also known as a split-plot ANOVA, is used to test for differences between two or more independent groups while subjecting participants to repeated measures. In this experiment, we study the impact of dementia on eTIV, nWBV and ASF while accounting for patients' two visits.

The eTIV's result is shown below.

	SS	F	p	np2
Group	37424.708	0.297	0.743	0.004
Visit	5573.920	0.003	0.003	0.061
Interaction	1004.783	0.438	0.438	0.012

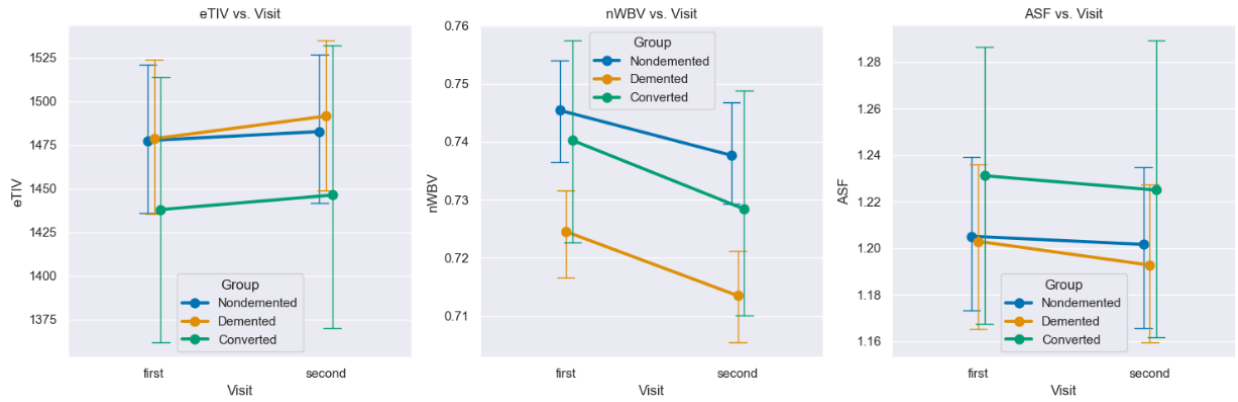
The nWBV's result is shown below.

	SS	F	p	np2
Group	0.034	6.712	0.002	0.087
Visit	0.007	94.251	0.000	0.401
Interaction	0.000	1.534	0.219	0.021

The ASF's result is shown below.

	SS	F	p	np2
Group	0.018	0.234	0.792	0.003
Visit	0.003	8.754	0.004	0.058
Interaction	0.001	1.028	0.361	0.014

Additionally, we plot the point plot of three variables grouped by visit and dementia.

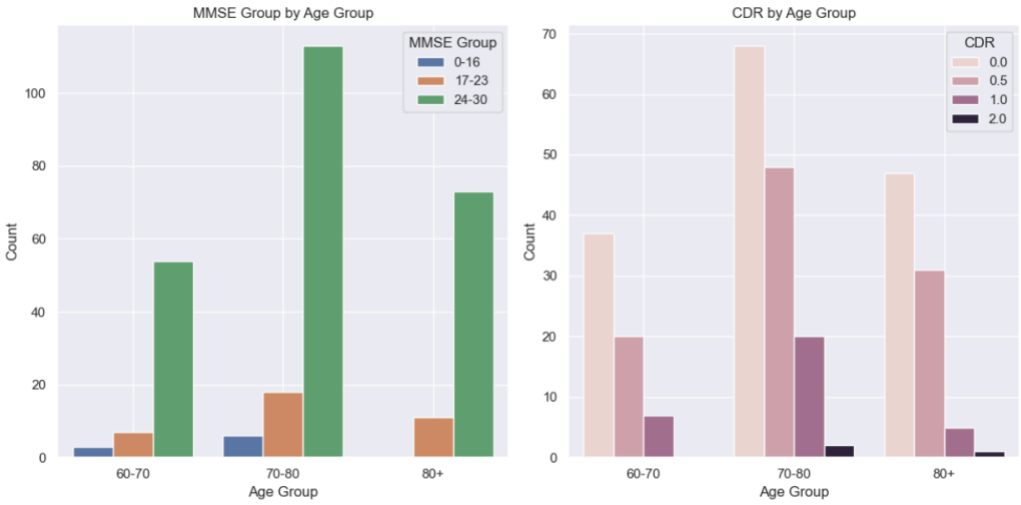


Conclusion: In summary, while dementia may not significantly influence eTIV and ASF when considering patients' two visits, it does have a significant impact on nWBV. This suggests that dementia status is associated with notable variations in normalized whole brain volume across multiple visits, highlighting potential structural differences or changes in brain volume related to dementia.

5. Age factor on MMSE and CDR

Research Question 2: What is the relationship between age and MMSE and CDR?

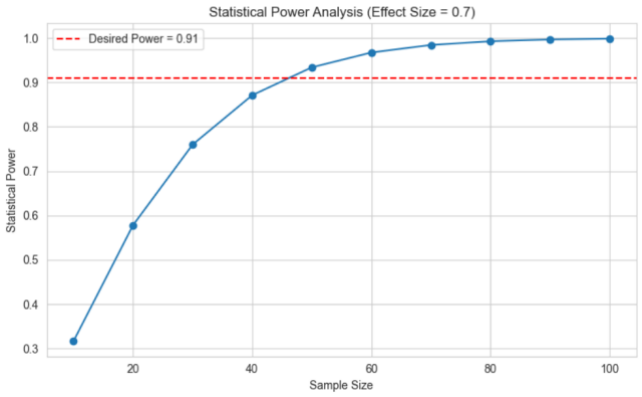
To explore this question, we visualize the countplot of MMSE group and CDR with age.



Based on the plot above and correlation map, we can conclude that age is not directly correlated with MMSE and CDR scores. However, older age is associated with an increased risk of more severe cognitive impairment.

6. Statistical power analysis

Statistical power analysis is a vital tool that helps researchers determine sample sizes effectively, evaluate the reliability of study results, and enhance the scientific validity and credibility of research. Larger sample sizes typically increase statistical power. Given power=0.91, alpha=0.05, effect_size=0.7, we can calculate the appropriate sample size is about 45. The power analysis plot is shown below,



7. Conclusion

Based on the comprehensive analysis, the study reveals that dementia does not significantly affect brain structure indicators eTIV and ASF, but it does have a pronounced impact on nWBV, indicating a potential link between dementia status and brain volume changes across multiple visits. Furthermore, there is no direct correlation observed between age and cognitive assessment metrics MMSE and CDR. However, with advancing age, there is an increased risk of more severe cognitive impairment.